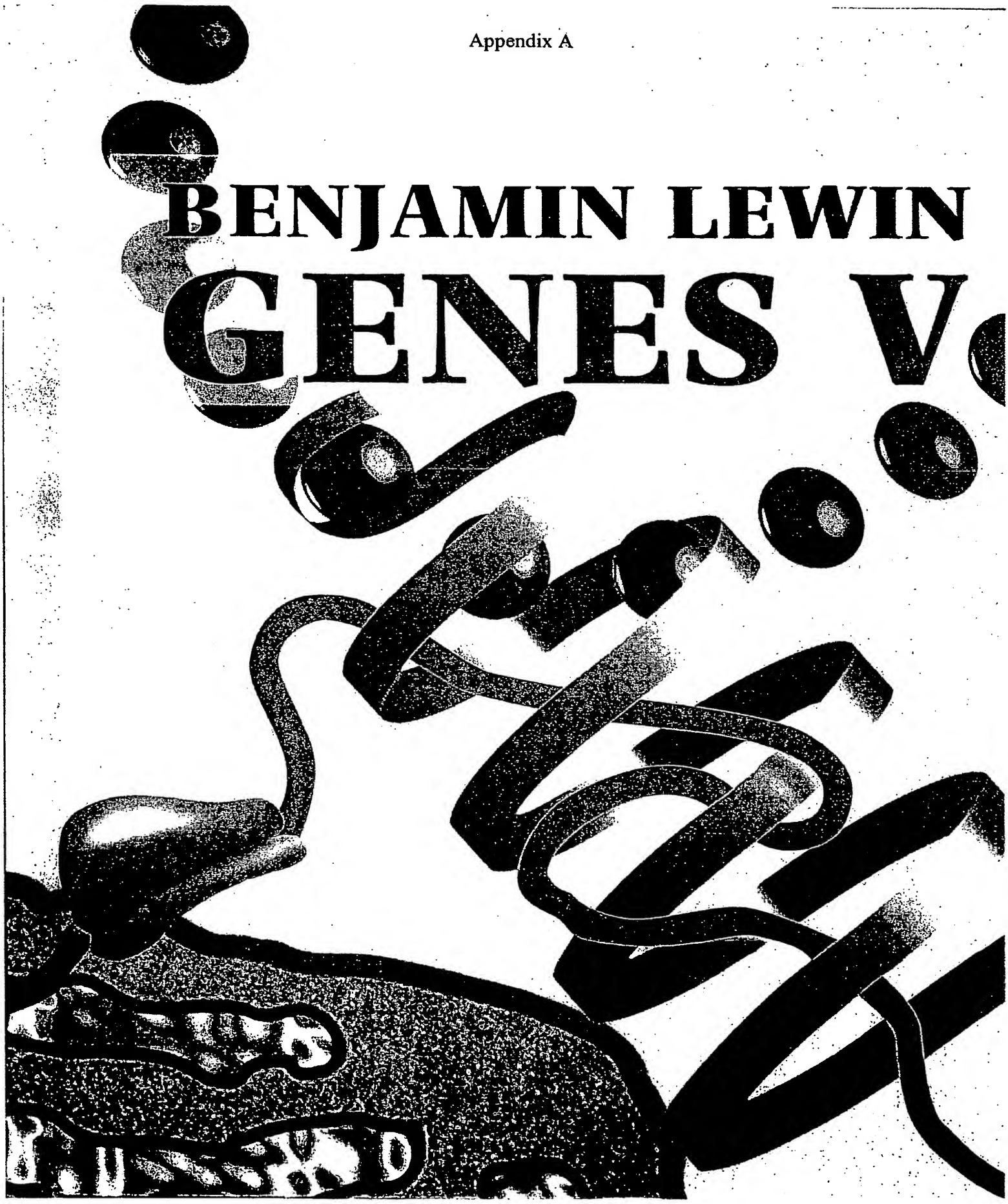


GENES V

Benjamin Lewin

OXFORD UNIVERSITY PRESS
Oxford New York Tokyo
1994

CH

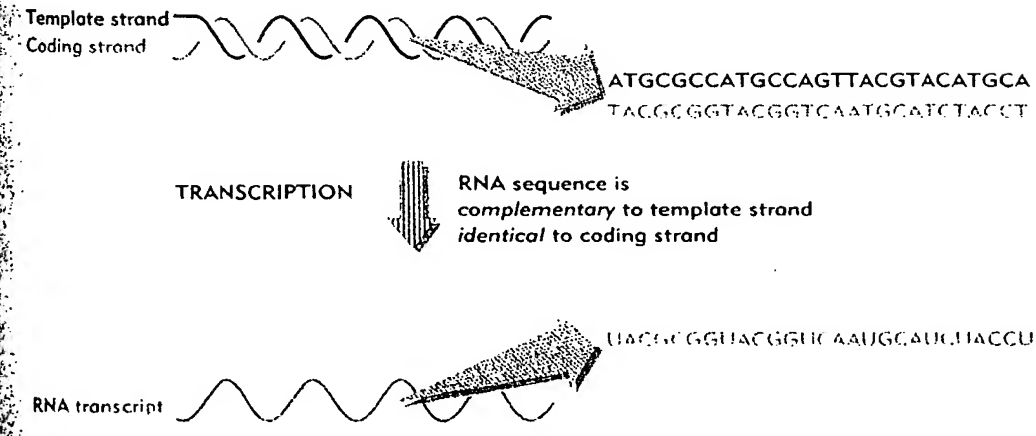
Control at initiation: RNA polymerase–promoter interactions

Transcription involves synthesis of an RNA chain representing one strand of a DNA duplex. By 'representing' we mean that the RNA is *identical in sequence* with one strand of the DNA, which is called the **coding strand**. It is *complementary* to the other strand, which provides the **template** for its synthesis. This relationship between double-stranded DNA and its single-stranded RNA transcript is recapitulated in Figure 14.1.

RNA synthesis is catalyzed by an enzyme, called **RNA polymerase**. Transcription starts when RNA polymerase binds to a special region, the **promoter**, at the start of the gene. The promoter surrounds the first base pair that is actually transcribed into RNA, the **startpoint**. From this point, RNA polymerase moves along the template, synthesizing RNA, until it reaches a **terminator** sequence. This action defines a **transcription unit** that extends from the

Figure 14.1

Overview: the function of RNA polymerase is to copy one strand of duplex DNA into RNA.



any additional material between them. A problem inherent in this technique is that there is no control over which pairs of blunt ends are joined together, so it is necessary first to perform the reaction and then to isolate the desired products from among the other products.

There are numerous variations of these methods. One technique uses short DNA duplexes ('linkers') that contain the *Eco*RI (or some equivalent) recognition palindrome. The linkers can be synthesized chemically, and are added covalently to the ends of a plasmid or an insert by blunt-end ligation. The inserted DNA can be retrieved by cleavage with *Eco*RI, but there are no restrictions on the original choice of sites to generate the ends. With sufficient manipulation, it is now possible to insert any foreign DNA fragment into any particular vector site, and to arrange for retrieval of the fragment when necessary.

When a foreign DNA fragment is inserted into a plasmid, it can be connected in either orientation, that is, with either of the ends of the foreign DNA joined to either of the ends of the plasmid. This does not matter when the purpose of cloning is simply to amplify the inserted sequence. However, it is important when the experiment is designed to obtain expression of the foreign DNA that requires insertion in a particular orientation.

In this case, populations of plasmids carrying the plasmid in either orientation are obtained by random insertion, after which they are characterized by restriction mapping to identify the desired class. Or the experiment is designed so as to permit insertion in one orientation only. For example, each of the DNAs, vector and insert, may be cleaved with *two* restriction enzymes that make different sticky ends, to generate the type of pattern where each DNA has the sequence

```
End 1 ——————————End 2
```

Now if only the two end-1 sequences can anneal together, and only the two end-2 sequences can anneal, the insertion can take place only in one orientation, generating the chimeric plasmid

```
End 1———Insert———End 2
  |                  |
End 1———Plasmid———End 2
```

Copying mRNA into cDNA

One of the principal uses of cloning technology is to isolate specific genes directly from the genome. Any particular gene represents only a very small part of a eukaryotic genome. In a typical mammal, the size of the genome is ~10^9 bp, so that a single gene of (say) 5000 bp represents only 0.0005% of the total nuclear DNA.

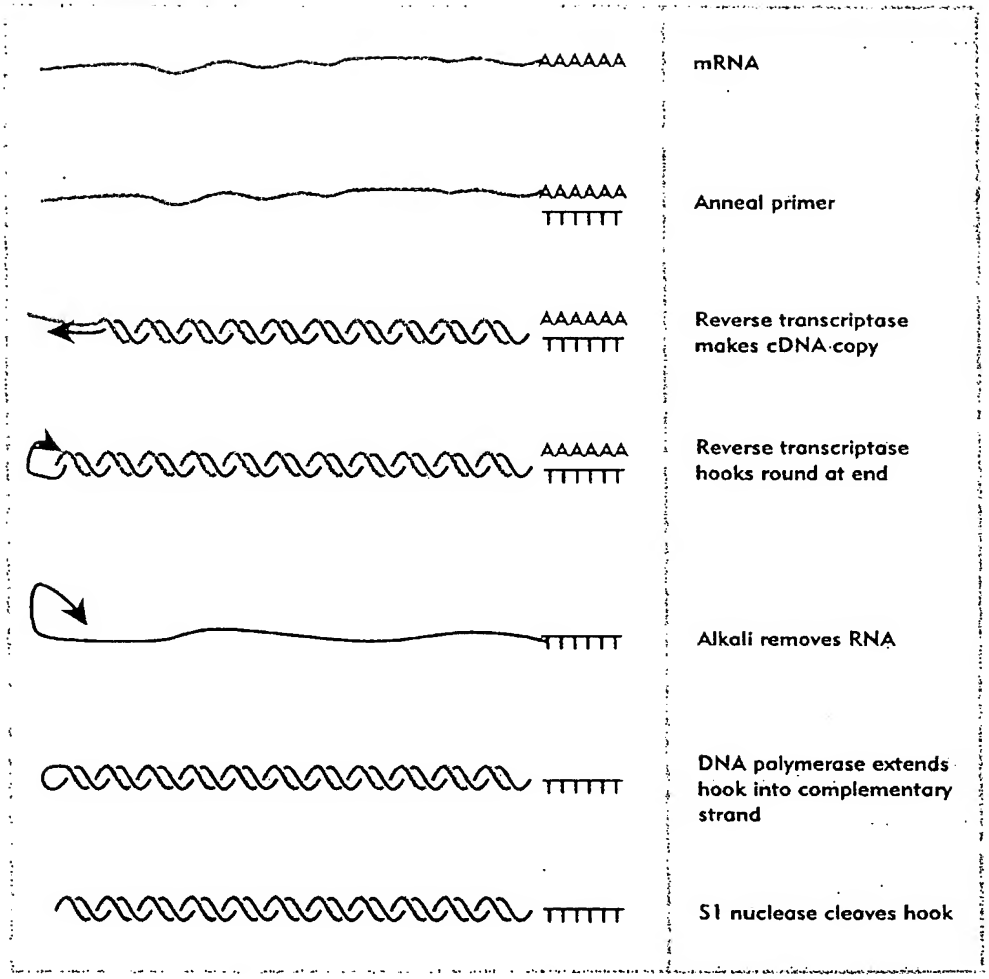
To identify such a tiny proportion, we need a very specific probe that reacts *only* with the particular sequence in which we are interested, to pick it out from the vast excess of other sequences. The usual technique is to use a highly labeled radioactive probe of RNA or DNA, whose hybridization with the gene is assayed by autoradiography.

For the purpose of obtaining a DNA sequence that represents a particular protein, the place to start is with mRNA, which, after all, is the template used to produce the protein *in vivo*. But it can be difficult to obtain the mRNA that represents a particular protein when the product is rare. There are several techniques for isolating an mRNA via the properties of its product, but a common problem is the requirement that the RNA must first be purified.

Rather than purify the RNA, a DNA copy of the RNA sequence is made. This has the two advantages that unlimited amounts of material can be obtained, and the DNA can be radioactively

Figure 21.5

mRNA can be copied into double-stranded DNA.



cleled, providing a much more powerful probe. The existence of reverse transcription makes it possible to synthesize a duplex DNA from any mRNA. This is especially easy for mRNAs that carry a poly(A) tail at the 3′ end, as illustrated in **Figure 21.5**. First, a primer is annealed to the poly(dA). It is a short sequence of oligo(dT), whose purpose is to provide a free 3′ end that can be used for extension by the enzyme reverse transcriptase. The enzyme engages in the usual 5′–3′ elongation, adding deoxynucleotides one at a time, as directed by complementary base pairing with the mRNA template. The product of the reaction is a hybrid molecule, consisting of a template RNA strand base-paired with the complementary DNA strand. The only practical problem is the propensity *in vitro* of reverse transcriptase to stop before it has reached the 5′ end of the mRNA. In this case, the resulting reverse transcript falls short of representing the entire mRNA, because it lacks some of the sequences complementary to the 5′ end. However, by judicious adjustment of the experimental conditions, usually it is possible to persuade reverse transcriptase to proceed all the way.

A useful reaction tends to occur at the end of the mRNA, where the enzyme causes the reverse

transcript to 'loop back' on itself, by using the last few bases of the reverse transcript as a template for synthesis of a complement. That is, the end of the complementary DNA is used to direct synthesis of a short sequence that is identical with the mRNA, and which displaces it. This creates a short hairpin, usually 10–20 bp long.

At this juncture, the original mRNA is degraded by treatment with alkali (a procedure that does not affect DNA). The product is a single-stranded DNA that is complementary to the mRNA; it is called cDNA.

The hairpin at the 3′ end of the cDNA provides a natural primer for the next step, the use of *E. coli* DNA polymerase I to convert the single-stranded cDNA into a duplex DNA via synthesis of the complementary strand. In this reaction, the enzyme uses the cDNA as template for synthesis of a sequence identical with the original mRNA. The product is a duplex molecule with a hairpin at one end. The hairpin is cut by the enzyme S1 nuclease (which specifically degrades single-stranded DNA) to generate a conventional DNA duplex.

The duplex DNA can be cloned to generate large amounts of a synthetic gene representing the mRNA sequence. This is called a **cDNA clone.** (From this terminology, a somewhat looser use of the term 'cDNA' has emerged, being taken to describe the duplex insert and not just the original single-stranded reverse transcript.)

The power of sequencing technology has made it possible to bypass most of the problems posed by rare mRNAs by targeting a probe directly to the mRNA sequence. One powerful technique requires knowledge of only a small sequence of the protein. Short oligonucleotides can be synthesized that correspond to this sequence. A variety of oligonucleotides can be made to cover possible alternative codons, especially at third base positions. These oligonucleotides can be used to isolate cDNAs or genomic DNA that include the sequence of the corresponding gene.

Isolating individual genes from the genome

The first step toward identifying the gene corresponding to a particular probe is to break the DNA of the genome into fragments of a manageable size. It is desirable to obtain the gene in as few fragments as possible (ideally only one). Usually the maximum lengths of DNA that can be manipulated directly are in the range of 15–20 kb. Sometimes it is not possible to obtain a gene in the form of a single fragment, and then its structure must be determined by piecing together the information gained from its various fragments. (We have discussed the use of overlapping fragments in Chapter 6.)

The best technique for fragmenting a genome is to make a restriction digest. Then every fragment ends in a site that was recognized by that particular enzyme. However, restriction sites may occur at inconvenient locations—for example, in the middle of a gene that is to be cloned. One way to avoid this is to use more than one restriction enzyme, that is to repeat the experiment with different enzymes whose recognition sites lie at different locations. But this is time consuming; and when a long sequence is involved, it may be difficult to find an enzyme that does not cleave within it.

When the DNA of an entire genome is digested with a restriction enzyme, the frequency of breakage is controlled by the length of the sequence recognized by the enzyme. The longer the sequence, the less often it occurs by chance. The probability that a particular 4 bp sequence will occur is $0.25^4 = 1/256$, so that an enzyme with such a short recognition sequence will cleave DNA rather frequently. The frequency declines to 1/1000 for a 5 bp sequence and to 1/4000 for a 6 bp sequence.

(This calculation assumes that each base